Archives of Biochemistry and Biophysics 644 (2018) 1–7

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Solution conformation of a cohesin module and its scaffoldin linker from a prototypical cellulosome

Albert Galera-Prat[a], David Pantoja-Uceda[b], Douglas V. Laurents[b,*], Mariano Carrión-Vázquez[a,**]

[a] Instituto Cajal, CSIC, Avda. Doctor Arce 37, E-28002, Madrid, Spain
[b] Instituto de Química Física "Rocasolano", CSIC, C/ Serrano 199, E-28006, Madrid, Spain

ARTICLE INFO

Keywords:
Scaffoldin
Biofuels
CipA
Climate change
Cellulosome
Cohesin
Linker
NMR spectroscopy

ABSTRACT

Bacterial cellulases are drawing increased attention as a means to obtain plentiful chemical feedstocks and fuels from renewable lignocellulosic biomass sources. Certain bacteria deploy a large extracellular multi-protein complex, called the cellulosome, to degrade cellulose. Scaffoldin, a key non-catalytic cellulosome component, is a large protein containing a cellulose-specific carbohydrate-binding module and several cohesin modules which bind and organize the hydrolytic enzymes. Despite the importance of the structure and protein/protein interactions of the cohesin module in the cellulosome, its structure in solution has remained unknown to date. Here, we report the backbone ^{1}H, ^{13}C and ^{15}N NMR assignments of the Cohesin module 5 from the highly stable and active cellulosome from *Clostridium thermocellum*. These data reveal that this module adopts a tightly packed, well folded and rigid structure in solution. Furthermore, since in scaffoldin, the cohesin modules are connected by linkers we have also characterized the conformation of a representative linker segment using NMR spectroscopy. Analysis of its chemical shift values revealed that this linker is rather stiff and tends to adopt extended conformations. This suggests that the scaffoldin linkers act to minimize interactions between cohesin modules. These results pave the way towards solution studies on cohesin/dockerin's fascinating dual-binding mode.

1. Introduction

Clostridium thermocellum is a cellulolytic, thermophillic and anaerobic bacterium inhabiting hot springs, soil and horse manure. This bacterium is among the most efficient microbes currently available for solubilizing lignocellulosic biomass and constitutes the prototypical cellulosome-producing bacterium [1,2]. Its main cellulosome complex, which is the best characterized cellulosomal system, is formed by a hierarchical assembly of different components. Here, a large modular protein, known as scaffoldin CipA is composed of nine type-I cohesin modules which afford anchorage to nine dockerin modules bearing enzymes. The presence of type-II dockerin in the CipA scaffoldin directs its binding to type-II cohesins of secondary scaffoldings. The presence of up to 7 of these cohesins in the same scaffoldin allows the assembly of complexes containing up to 63 enzymes. The high cellulolytic activity and conformational stability as well as the ethanolic anaerobic glycolysis metabolism of *C. thermocellum* have attracted significant interest to this system, and it is hoped that an improved understanding of the structure and interactions of the *C. thermocellum* cellulosome will facilitate the design of improved cellulosomes for industrial applications such as the production of biofuels [1].

Due to their importance, the structure of type-I cohesin modules has already been investigated by X-ray crystallography [3–5]. These studies revealed a "jelly-roll" type fold with nine β-strands arranged with mostly anti-parallel topologies organized into a β-barrel with two principal β-sheets. A large and well ordered hydrophobic core, composed of numerous aromatic and aliphatic sidechains, many of the latter being β-branched, packs the space between the two β sheets. Increased hydrophobicity and sidechain rigidity, which are afforded by aromatic residues and β-branched residues, are well-known hallmarks of thermophilic proteins [6].

These cohesin modules are linked to each other in scaffoldin by Thr/Pro-rich linkers containing between 3 and 721 residues [7]. Despite the importance of these linkers in the conformation, dynamics and even activity of the cellulosome [8], high resolution structural information on them is still limited. A short linker between two cohesin modules was solved by X-ray crystallography revealing contacts between the cohesin and the linker [9,10] and even between two adjacent cohesin modules [10]. Additionally, part of a longer linker of the same scaffoldin was also crystallized [9]. Nevertheless, the former is a rather

* Corresponding author.
** Corresponding author.
E-mail addresses: dlaurents@iqfr.csic.es (D.V. Laurents), mcarrion@cajal.csic.es (M. Carrión-Vázquez).

https://doi.org/10.1016/j.abb.2018.02.016
Received 27 October 2017; Received in revised form 20 February 2018; Accepted 23 February 2018
Available online 24 February 2018
0003-9861/ © 2018 Elsevier Inc. All rights reserved.

short linker, especially among complex cellulosomes while the latter showed elevated temperature factors and distinct conformations in two different crystals. Thus, the formation of stable structures that can be crystallized may not represent the general behavior but a special case, as the authors of these studies have indicated. On the other hand, several SAXS studies, using both native [11,12] or artificial [13,14] sequences, support the idea that the linkers between cohesins are the regions that afford flexibility to the cellulosome. Using electron microscopy, a CipA fragment containing cohesin 3 to 5 was found to adopt a mostly defined structure [15], although the resolution of this negative staining study was too low to define the conformation of the linkers.

Cohesin modules possess an elevated mechanical stability, which is especially high for internal modules of the cellulosome (*i.e.* those connecting the cell to the cellulose substrate) [16], such as module 5 studied here. Diverse modular interactions between the cohesin modules and dockerin-linked enzymes build up potent hydrolytic complexes. The interaction between cohesin and dockerin is relatively strong with the force required to dissociate cohesin from dockerin ranging from about that of half of a covalent bond [17] to below 100 pN [18]; the latter value was recently measured directly using a novel single molecule force spectrometry strategy.

Whereas many details of the cohesin-dockerin interaction have been revealed by X-ray crystallography [19], some features of this complex could well be conditioned by crystal packing forces. This is particularly important considering that several type-I dockerins contain two very similar potential binding modes to cohesin, a truly remarkable and unique feature of this interaction. Both have been seen to be functional on the basis of mutational studies [20], yet only one appears to be occupied in the crystal [21]. The results reported here should clear the way towards the resolution of this important issue by future studies of the cohesin-dockerin interaction in solution.

To address these questions, we have used NMR methods to study the conformation in solution of the fifth type-I cohesin module and the adjacent linker segment from the *Clostridium thermocellum* CipA scaffoldin (CtCoh5) in solution.

2. Materials and methods

2.1. Expression and purification of CtCoh5

The fifth cohesin module from the *Clostridium thermocellum* cellulosome scaffoldin CipA (CtCoh5) was produced by recombinant expression methods in *E. coli* (strain: C41(DE3)) from a plasmid constructed *ad hoc* for this study containing a His-tag, the TEV recognition site followed by two Gly residues and the CtCoh5 sequence as annotated in UniProt. Isotopic labelling with ^{13}C and ^{15}N was performed by growing cells in M9 minimal media containing $^{15}NH_4Cl$ and ^{13}C-glucose as the exclusive sources of nitrogen and carbon, respectively, following a published protocol [22]. The protein was purified using Ni^{2+}-affinity chromatography using a nickel-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare). Samples containing the protein were concentrated into pH 8 buffer containing 50 mM Tris, 150 mM NaCl, 1 mM DTT. The His-tag of the protein was cleaved by incubating the protein sample with TEV protease overnight at 37 °C. The sample was then purified again using Ni^{2+}-affinity chromatography and the cleaved protein was recovered from the unbound fraction, and concentrated into 2 mM ammonium bicarbonate solution before being lyophilized.

2.2. Linker peptide

A linker peptide with the sequence GDTTEPATPTTPVTTPTTTDDLDA, which appears four times between cohesins of *C. thermocellum* CipA scaffoldin, and in particular flanking cohesin 5 (Sup. Fig. 1), was obtained commercially from GenScript. Its purity was over 95% as judged by HPLC and its identity was confirmed by mass spectrometry and NMR spectroscopy.

2.3. Assignments

A standard suite of NMR spectra 2D [^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC]; 3D [^{1}H-^{15}N HSQC·NOESY, 3D HNCO, HNcaCO, HNCAi, HNcoCA, HNcoCACB, CBCAcoNH] [23] and as well as the 2D BECE [24] experiment, which provides amino acid type identification of the NH correlation signals, were recorded for the CtCoh5 module on a Bruker AV 800 MHz (^{1}H) spectrometer equipped with Z-gradients and a cryoprobe at 25 °C. The CtCoh5 concentration was 1.26 mM and the solution contained 90% milliQ water and 10% D_2O (Cambridge Isotope Labs), 8.5 mM KH_2PO_4, 1.5 mM K_2HPO_4, 0.7 mM NaN_3 (pH 6.1) and 0.05 mM sodium 4,4-dimethyl-4-sila-pentase-1-sulfonate (DSS) as the internal chemical shift reference. Spectra were transformed and processed using Bruker TOPSPIN 2.1 software and manually assigned with the aid of SPARKY [25]. As an independent corroboration, the assignments were also assigned automatically using an in-house program based on the MARS algorithm [26] with novel assignment correction capacities. Both sets of assignments were very similar and the few cases for which discrepancies arose were resolved by manual inspection of the spectra.

For the linker peptide, NMR spectra: 2D ^{1}H-^{1}H NOESY (mixing time = 150 ms, 2D ^{1}H-^{13}C HSQC, 2D ^{1}H-^{1}H TOCSY (mixing time = 60 ms), 2D ^{1}H-^{1}H COSY (mixing time = 80 ms) and 2D ^{1}H-^{15}N HSQC (at natural abundance ^{15}N) spectra were recorded on a Bruker Avance spectrometer operating at 600 MHz (^{1}H) and equipped with a cryoprobe and Z-gradients in aqueous solution containing the same composition as listed above for the CtCoh5 studies. A ^{1}H-^{13}C HSQC spectrum was acquired at natural abundance ^{13}C in 99.9% D_2O containing the same buffer. The peptide's concentration was 2.8 mM. Spectra were recorded at lower temperature (5 °C) to enhance structure formation.

2.4. Data analysis

Populations of partial secondary structure were calculated for the linker peptide utilizing the program δ2D [27]. The difference between the chemical shift values of the beta and gamma carbons of Pro ($\delta\ ^{13}C\beta - \delta\ ^{13}C\gamma$) was used to determine whether Xaa-Pro peptide bonds are in the *cis* or *trans* configuration, as described previously [28]. NOE crosspeaks were quantified using SPARKY as well as Bruker TopSpin 4.0.2 and consistent values were obtained using both programs. The program TALOS+ by Shen et al. [29], was used to analyze the experimental chemical shift values obtained here to identify elements of secondary structure in the folded CtCOh5 module and to obtain torsion angle restrictions for the linker peptide. TALOS+ also applies the method of Berjanski & Wishart [30] to predict the backbone flexibility based on experimental chemical shift data. The torsion angle restrictions were employed to calculate a family of conformers for the linker peptide using the program CYANA [31]. Finally, the programs MOLMOL [32] and KiNG [33] were then utilized to visualize and compare these structures and to prepare figures.

2.5. Homology modeling

The program package SWISS MODEL [34] was used to calculate a structural model based on the close sequence similarity of CtCoh5 to other cohesin modules of known structure. This approach is justified by the closely similar secondary structure of CtCoh5, as determined here by analysis of the experimental chemical shifts, to cohesin modules whose 3D structures had been previously elucidated by X-ray crystallography.

3. Results

On the basis of the obtained NMR spectra, it was possible to assign 98.6% of ^{13}CO and $^{13}C\alpha$ resonances, 97.1% of ^{1}HN and ^{15}N signals and

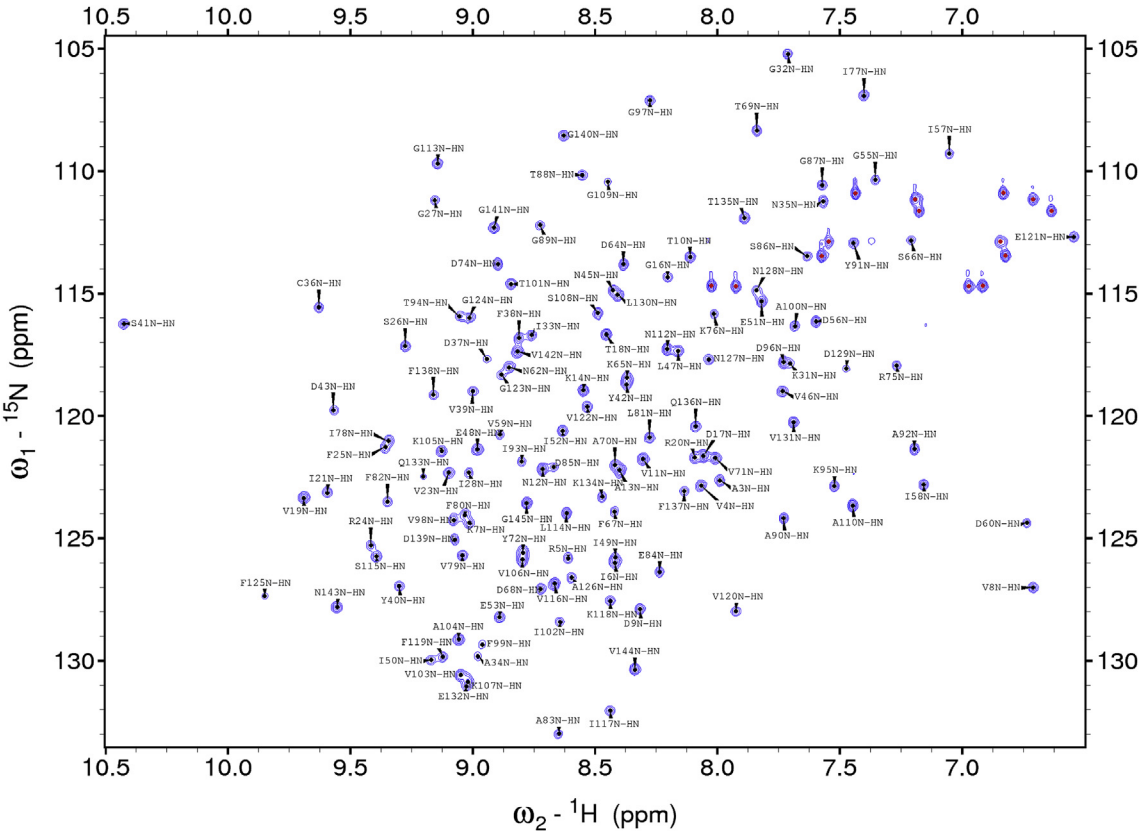

Fig. 1. ^{1}H-^{15}N HSQC Spectrum of CtCoh5 recorded at pH 6.0, 25 °C.
The CtCoh5 sample contained 90% H_2O, 10% D_2O, 10 mM potassium hydrogen phosphate buffer, 0.7 mM sodium azide and 0.050 mM DSS as the chemical shift reference. Backbone amide ^{1}H-^{15}N resonances are labeled. Asn and Gln side chain amide resonances, $-NH_2$, are marked with a red dots. Note that this spectrum does not include signals from the linker, which was studied as an autonomous peptide. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

96.9% of ^{13}C*β* resonances of CtCoh5. Most of the missing signals correspond to residues Gly 1 and Gly 2 at the N-terminus and Pro 29. Regarding the linker, 95.8%, 91.3%, 94.7% and 100% of ^{13}C*α*, ^{13}C*β*, ^{1}HN and ^{1}H*α* resonances were assigned. The ^{13}CO, ^{13}C*α*, ^{13}C*β*, ^{1}HN and ^{15}N chemical shift values of the folded CtCoh5 module and the ^{13}C*α*, ^{13}C*β*, ^{1}HN and ^{1}H*α* chemical shift modules of the linker have been deposited in the **BMRB** under access code **27180**. The assigned 2D ^{1}H-^{15}N HSQC spectrum of CtCoh5, obtained at 25 °C in 8.5 mM KH_2PO_4, 1.5 mM K_2HPO_4, 0.7 mM NaN_3 (pH 6.1) is shown in Fig. 1. The spectrum shows a high chemical shift dispersion that is characteristic of well-folded proteins. Notably, the ^{1}HN signal of Ser 41, a residue that is sandwiched between two Tyr residues, is shifted far down field to 10.40 ppm. Well-folded proteins generally have one or more turn or loop segments whose ^{1}H-^{15}N HSQC crosspeaks are broad or missing due to conformational sampling on the microsecond - millisecond time scale. It is remarkable that CtCoh5's ^{1}H-^{15}N signals' intensities and line widths are similar; there are no broad signals which are characteristic of multiple conformers interconverting on 10^{-4} – 10^{-2} s time scales. Moreover, no sharp signals are observed between 8.0 and 8.7 ppm ^{1}H (which are typical of disordered segments, see for example: Mompeán et al. [35]) and this observation rules out the presence of unfolded segments. The ^{13}C*β* and ^{13}Cγ chemical shifts of proline residues are faithful indicators of the *cis/trans* isomeric state of Xaa-Pro peptide bonds. An analysis of these values reveals that eight Xaa-Pro peptide bonds are in the *trans* conformation (Table 1), which is the most stable configuration.

The determined chemical shift values were analyzed using TALOS + to identify the type and position of elements of secondary structure. Nine *β*-strands, connected by relatively short turns, loops or short helical elements were detected. The results for the folded CtCoh5 module are summarized and compared to module 7 of CtCoh, whose structure had been resolved by X-ray methods (PDB 1AOH) [36], in Fig. 2. The close agreement found indicates that the crystal structure is preserved in solution. Encouraged by this finding, and the high level of sequence identity among cohesin modules (Sup. Fig. 1A), we calculated a structural model for CtCoh5 by homology modeling. Using the SWISS MODEL package [34], four cohesin or cohesin-like module structures, previously elucidated by X-ray crystallography, with significant sequence identity to CtCoh5 were identified (Table 2). Their respective 3D structures were utilized as the basis to calculate four structural models of CtCoh5. Of these, the model built using CtCoh7 (PDB 1AOH) had the highest global quality scores and no local quality scores below 0.6 (Table 2), which are characteristics of a very accurate and precise model [34]. This structure, shown in Fig. 3, consists of a *β*-barrel composed of long *β*-strands linked by turns, short loops and one turn of helix composed of residues P63 to F67. The structural model features a

Table 1
Isomeric State of Xaa-Pro peptide bonds in CtCoh5.

Residue	δ^{13}Cβ (ppm)	δ^{13}Cγ (ppm)	Δ^{13}Cβ − ^{13}Cγ (ppm)	Xaa-Pro isomeric conformer[a]
Pro 15	31.64	28.76	2.88	*TRANS*
Pro 22	33.68	28.41	5.27	*TRANS*
Pro 44	31.89	26.45	5.44	*TRANS*
Pro 54	31.27	26.94	4.32	*TRANS*
Pro 61	32.06	27.02	5.04	*TRANS*
Pro 63	32.06	27.29	4.77	*TRANS*
Pro 73	32.24	27.45	4.79	*TRANS*
Pro 111	32.11	26.77	5.33	*TRANS*

[a] Based on the average difference of 4.5 ± 1.2 ppm for *trans* conformers and 9.6 ± 1.3 ppm for *cis* conformers reported by Schubert et al. (2002) [28].

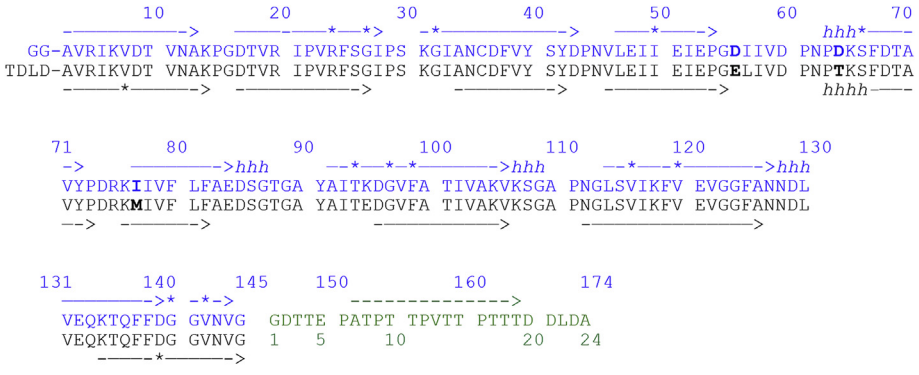

Fig. 2. The Secondary Structure of Cohesin Module 5 (solution NMR) is similar to that of Cohesin Module 7 (X-ray crystallography).
The amino acid sequence of CtCoh5 is shown in blue for the folded module and green for the linker. The sequence of CtCoh7 appears in black. Distinct residues are shown in bold. β-strands are represented as arrows and α-helical conformations are indicated by the letter "h" in italics. The green broken arrow for the linker represents extended (β-strand plus PPII conformers). Asterisks indicate points of curvature in the β strands of X-ray structure, or weakly predicted β conformations in the TALOS+ analysis of the NMR data. β-strands and α-helices are detected from the TALOS+ data as three or more residues with good values of ϕ,φ space corresponding to β-strand or α-helix in the Ramachandran map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

substantial hydrophobic core, which is dominated at its widest point by seven Phe residues.

Chemical shift dispersion is a reliable indicator of protein backbone dynamics [30]. On a scale where an order parameter value of one indicates complete rigidity and a value of zero represents total flexibility, the folded cohesin module revealed high predicted order parameters, approaching 0.9, for the elements of secondary structure upon analysis of their chemical shift values (Fig. 4). Somewhat lower values, ranging from 0.7 to 0.8, were obtained for the turn and loop segments connecting elements of secondary structure. Interestingly, for the center of the linker, the predicted ordered parameters range from 0.6 to 0.7; this is considerably more rigid than a flexible random coil. This intriguing result led us to examine the linker's conformation in more detail.

The linker peptide's sequence GDTTEPATPTTPVTTPTTTDDLDA corresponds to that of the consensus linker in *C. thermocellum* scaffoldin (Sup. Fig. 1B). In contrast to the cohesin module, the linker's ^{1}HN resonances show minimal chemical shift dispersion, which is a hallmark of disordered polypeptides (Sup. Fig. 2). Moreover, like intrinsically disordered proteins, the linker has a low content of hydrophobic and aromatic residues and high content of proline (4 residues), anionic (D +E = 5), and polar residues (T = 10). However, this rich content of Pro and β-branched Thr will also disfavor helical conformations [37] and favor extended ones [38].

Intense Pro Cβ and Cγ signals where found at the chemical shift values corresponding to the *trans* conformer, which indicates that the linker's E_5-P_6, T_8-P_9, T_{11}-P_{12} and T_{15}-P_{16} peptide bonds are all predominantly *trans*. Nevertheless, a pair of weak crosspeaks in the 2D ^{1}H-^{13}C HSQC at 1.96 (^{1}H) & 24.8 (^{13}C) ppm and 1.84 & 24.8 ppm could arise from Pro ^{13}C-^{1}Hγ signals when Xaa-Pro peptide bonds in the *cis* conformation, and a second pair of faint crosspeaks at 2.03 & 29.8 ppm and 1.80 & 29.8 ppm may be attributed to Glu 5's ^{1}H-^{13}C β when Glu5-Pro6 is *trans* (Sup. Fig. 3). Although these signals could not be corroborated in the 2D ^{1}H-^{1}H COSY, NOESY and TOCSY spectra, integration shows that they are about 10% as strong as the intense *trans* resonances. This is approximately the proportion of *cis* conformers populated in unstructured short peptides containing Thr-Pro (9.4%) and Glu-Pro (9.0%) [39]. Therefore, the presence of minor populations of *cis* Xaa-Pro peptide bonds can not be ruled out. The tendency of the linker peptide to adopt extended conformations is probably not strong enough to lock Xaa-Pro peptides bonds exclusively in the *trans* conformation, as sometimes occurs for peptides with a well defined structure [40].

When its assigned chemical shifts values were analyzed with the program δ2D [27], a substantial population of extended conformers was detected (Fig. 5). In fact, for residues 8 to 17 in the center of the linker, considerable populations of extended conformations including β-strands (10%) and polyproline-II helices (24%) are detected. Throughout the whole linker, the content of helix is negligible (< 2%), except at the C-terminus. NOE crosspeak intensities provide experimental information on secondary structure. An evaluation of the intensities of non-overlapped Hα_{i-1}-HN$_i$ and Hα_i-HN$_i$ NOE resonances revealed that the former tend to be more intense (Table 3). This is consistent with low population of helical conformations and a predominance of extended conformers since the Hα_{i-1}-HN$_i$ distance is shorter (2.2 Å) than the Hα_i-HN$_i$ (2.8 Å) distance in a β-strand [41]. The opposite is true for an α-helix where the Hα_{i-1}-HN$_i$ distance is longer (3.5 Å) than that of Hα_i-HN$_i$ (2.8 Å). Furthermore, only a few weak HN$_i$ – HN$_{i+1}$ NOE crosspeaks are observed, which arise from residues near the peptide's C-terminus (Table 3). These signals are generally abundant in α-helices where the HN$_i$ – HN$_{i+1}$ distance is short (2.8 Å) but are weak or absent in extended conformations like the β-strand where these nuclei are further apart (4.2 Å) [41].

To further test this finding, we also analyzed the chemical shift data using the TALOS+ program. The results confirm the absence of helical structure and corroborate the presence of extended β-strand and polyproline-II conformers for residues 7–19. The TALOS+ analysis yielded good torsional angle constraints for most residues (Table 4). Using these constraints, a family of backbone structures was calculated using the program CYANA. For the well-defined central region, the twelve

Table 2
Structural Models for the CtCoh5 Module Obtained using SWISS MODEL [34].

Protein	PDB file	Resol-ution (Å)[a]	% SI[b]	Ligand[c]	GMQE	Qmean	Local Quality Score
Cellulosome-Integrating protein CipA (cohesin module 7)	1AOH.1.A	1.7	96.5	none	0.99	1.95	All > 0.6
Cellulosome-Integrating protein CipA (cohesin module 2)	1OHZ.1.A	2.2	74.65	6 Ca^{++}	0.89	1.16	All > 0.6
Cellulosome-Integrating protein CipA (cohesin module 2)	2CCL.2.A	2.0	74.65	2 Ca^{++}	0.87	0.78	All > 0.6 except for a few residues at the C-terminus
Cellulosome-Integrating protein CipA (cohesin module 2)	1ANU.1.A	2.1	76.09	None	0.86	0.50	All > 0.6 except for a few residues at the C-terminus and residues 15 - 20

[a] Resolution (A) of the PDB structure solved by X-ray crystallography.
[b] Percent Sequence Identity (% SI) to CtCoh5.
[c] The ligand (*if any*) that is bound to the protein structure used to model CtCoh5.

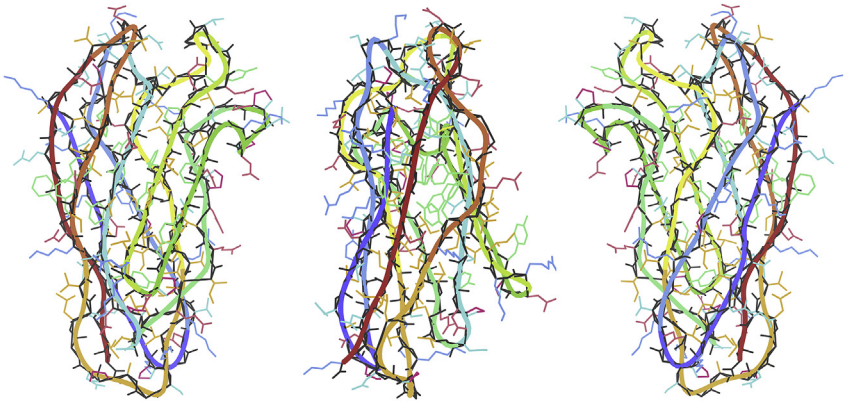

Fig. 3. Atomistic structural model of CtCoh5.
The three views of CtCoh5 were obtained by rotating the central figure clockwise or counter-clockwise by 90° about the z-axis to obtain the left and right views. The mainchain is shown in black and sidechains are colored: blue = cationic, sea green = polar; red = anionic and Pro, green = aromatic, orange = aliphatic. The Cα trace is represented as thick line colored with rainbow color gradient from blue at the N-terminus to red at the C-terminus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

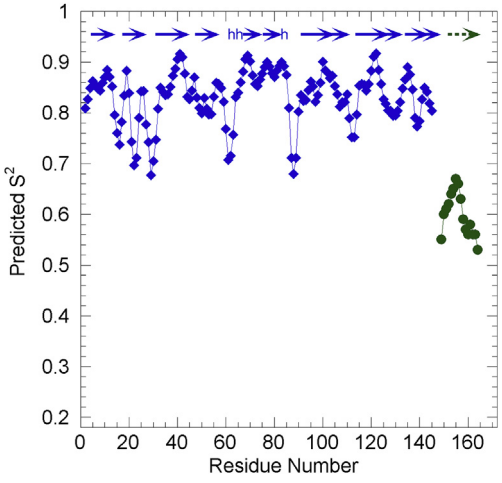

Fig. 4. Predicted backbone dynamics.
The predicted order parameter (S^2) which ranges from 1.00 for completely rigid peptides to 0.00 for totally flexible ones is plotted in blue for the CtCoh5 domain and green for the linker peptide. The approximate position of elements of secondary structure is indicated by blue arrows for β-strands, "h" for helix and a dotted green arrow for extended conformers (β-strands and polyProline II helixes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

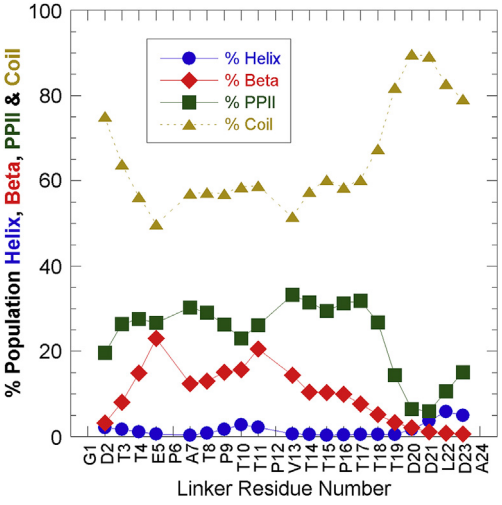

Fig. 5. The linker segment adopts some extended conformers and no α-helix.
The percent population of elements α-helical (blue circles), β-stands (red diamonds) and polyPro II helices (green squares) secondary structure as well as coil (golden diamonds). These results are based on the analysis of the chemical shift values using the program δ2D [27], as described in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
NOEs crosspeaks intensities.[a]

Crosspeak	Intra HA_i-HN_i	$HA_i - HN_{i+1}$	$HN_i - HN_{i+1,2,3}$
G1HA-D2NH		157: **8.81**	
D2HA-D2NH	98: **2.23**		
D2HA-T3HN		192: **12.8**	
T4HA-E5NH		457: **34.4**	
T4HA-T4NH	455: **55.9**		
P6HA-A7HN		851: **50.1**	
T8HA-T8HN	258: **43.6**		
P9HA-T10HN		859: **64.4**	
T11HA-HN	113: **8.93**		
P12HA-V13HN		993: **70.3**	
V13HA-HN	126: **12.7**		
V13HA-T14HN		1090: **80.3**	
T14HA-T15HN		778: **51.8**	
P16HA-T17HN		762: **58.7**	
T18 HN-D20HN			126: **3.42**
D20 HN-T18HN			77: **1.93**
D23HA-A24HN		98: **5.64**	
L22HN-T19HN			69: **9.78**
A24HA-A24HN	21: **0.96**		

[a] Values in plain text are peak intensities calculated with Sparky 3.113 and values in **bold** are peak integrals calculated with Bruker TopSpin 4.0.2. The values obtained from these two different programs are strongly correlated (R > 0.96).

Table 4
Torsion angle restrictions for calculation of the linker backbone conformation.

Residue	Phi lower limit	Phi upper limit	Psi lower limit	Psi upper limit
Asp 2	−153.8	−13.8	75.9	215.9
Thr 3	−142.3	−58.8	105.1	163.4
Glu 5	−105.9	−52.4	113.9	166.8
Ala 7	−125.5	−59.0	81.3	170.6
Thr 8	−141.7	−56.4	67.8	167.6
Pro 9	−83.7	−41.7	128.1	168.1
Thr 10	−143.8	−52.2	110.5	167.2
Thr 11	−157.1	−42.7	55.9	175.9
Pro 12	−83.9	−43.9	129.4	169.4
Val 13	−140.9	−55.5	109.4	161.9
Thr 14	−152.2	−45.6	104.2	164.4
Thr 15	−156.5	−32.6	50.4	174.2
Thr 17	−162.5	−59.8	84.2	162.8
Thr 18	−146.8	−69.1	89.1	164.9
Thr 19	−123.3	−76.5	106.5	146.5

conformers with the lowest CYANA target energy function values show considerable structural similarity (RMSD = 2.8 Å) (Fig. 6). Additional structural parameters are reported in Sup. Table 1. The extended nature of the linker is clearly evident.

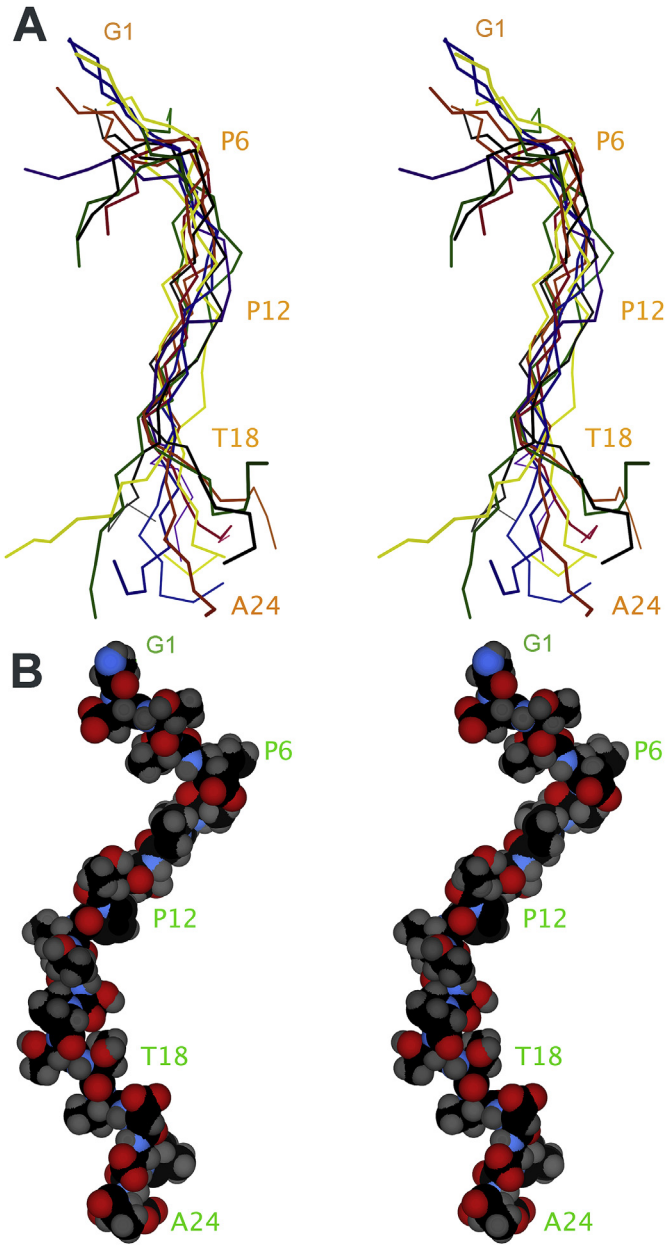

Fig. 6. The interdomain linker adopts preferred extended conformations.
A. The C*α* backbone trace of the twelve linker conformers with the lowest CYANA target functions are represented in cross-eyed stereo in different colors. The N-, C-termini and every sixth residue are labeled. For the central region (residues 3–19) the backbone RMSD of these twelve structures is 2.8 Å.
B. All atom, van der Waals, cross-eyed stereo representation of the linker conformer with the lowest CYANA target function. Atoms are colored as follows: hydrogen = gray, carbon = black, nitrogen = blue and oxygen = red. The N-, C-termini and every sixth residue are labeled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Although some high resolution structures for cohesin or cohesin-like modules have been determined by X-ray crystallography, prior to this study only limited atomic-resolution data had been published on its conformation in solution [42,43]. The results reported here show that the CtCoh5 adopts the classical cohesin fold in solution, where there are no crystal packing interactions which might perturb its conformation. Considering the high level of sequence identity among the *Clostridium thermocellum* cohesin modules (Sup. Fig. 1A), they can all be expected to adopt similar structures in solution.

Despite years of study, there are still important doubts concerning the interactions between cohesin modules and the modules they bind to. In particular, data are available which strongly suggest that dockerin, a small all-helical module with an interesting internal symmetry, can bind using either one of two quasi-equivalent loops [20]. The chemical shift data reported here for CtCoh5 and deposited in the biological magnetic resonance data base (**BMRB** under access code **27180**) pave the way for future studies of cohesin/dockerin interactions in aqueous solution.

Despite its rich content of polar residues and proline, which are common in intrinsically disordered proteins, the linker segment shows a significant tendency to avoid helical conformations and prefers to adopt extended structures such the polyproline-II helical conformation and *β*-strands. This is not so surprising considering that the linker segment is quite rich in Thr residues, which disfavor helical conformations [37] but has a strong intrinsic propensity to adopt *β*-conformation in internal *β*-strands [44], edge *β*-strands [45] as well as exposed positions in *β*-sheet [38]. The high content of negatively charged residues (5 out of 24 residues) will produce electrostatic repulsion which will also tend to expand the chain [46]. In addition, the glycosylation of Thr residue(s) [47,48] in the linker segment is expected to further favor extended peptide backbone conformations. Based on the rather stiff nature of this linker segment and its tendency to adopt extended conformations, it is improbable that it would interact significantly with the cohesin modules and will rather tend to space out the cohesin modules, so as to minimize interactions between them. The other *Clostridium thermocellum* linker segments have identical or similar sequences (Sup. Fig. 1B) and will therefore likely adopt similar extended conformers which should act to separate the cohesin modules.

Conflicts of interest

The authors declare that they have no conflict of interest.

Ethical standards

All experiments reported here are in compliance with the laws of Spain and the European Union.

Acknowledgements

This work was supported by grants from a Seventh Framework Programme in Nanosciences, Nanotechnologies, Materials & New Production Technologies (7PM -NMP 2013-17, 604530-2) and the ERA-IB-ERANET-2013-16 (EIB.12.022) through the Spanish MINECO (PCIN-2013-011-C02-01) to MC-V. and from grants SAF-2016-76678-C2-2-R (DVL) & CTQ-2014-52633-P (DPU). AGP acknowledges financial support from an FPU fellowship from Spanish MECD.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.abb.2018.02.016.

References

[1] L. Artzi, E.A. Bayer, S. Moraïs, Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides, Nat. Rev. Microbiol. 15 (2016) 89–95.
[2] E.A. Bayer, Y. Shoham, R. Lamed, Lignocellulose-decomposing bacteria and their enzyme systems, in: E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, F. Thomson (Eds.), The Prokayotes. Prokayotic Physiology and Biochemistry, Springer, Berlin, 2013, pp. 215–265.
[3] L.J.W. Shimon, E.A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, F. Frolow, A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly, Structure 5 (1997) 381–390.
[4] G.A. Tavares, P. Béguin, P.M. Alzari, The crystal structure of a type I cohesin domain at 1.7 Å resolution, J. Mol. Biol. 273 (1997) 701–713.
[5] S. Spinelli, H.P. Fiérobe, A. Belaïch, J.P. Belaïch, B. Henrissat, C. Cambillau, Crystal

structure of a cohesin module from *Clostridium cellulolyticum*: implications for dockerin recognition, J. Mol. Biol. 304 (2000) 189–200.
[6] W.F. Li, X.X. Zhou, P. Lu, Structural features of thermozymes, Biotechnol. Adv. 23 (2005) 271–281.
[7] E. Bayer, S. Smith, I. Noach, O. Alber, J. Adams, R. Lamed, L. Shimon, F. Frolow, Can we crystallize a cellulosome? Biotechnology of Lignocellulose Degradation and Biomass Utilization, 2009, pp. 183–205.
[8] Y. Vazana, Y. Barak, T. Unger, Y. Peleg, M. Shamshoum, T. Ben-Yehezkel, Y. Mazor, E. Shapiro, R. Lamed, E.A. Bayer, A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates, Biotechnol. Biofuels 16 (6) (2013) 182.
[9] I. Noach, F. Frolow, O. Alber, R. Lamed, L.J.W. Shimon, E.A. Bayer, Intermodular linker flexibility revealed from crystal structures of adjacent cellulosomal cohesins of *Acetivibrio cellulolyticus*, J. Mol. Biol. 391 (2009) 86–97.
[10] I. Noach, M. Levy-Assaraf, R. Lamed, L.J.W. Shimon, F. Frolow, E.A. Bayer, Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad, J. Mol. Biol. 399 (2010) 294–305.
[11] M.A. Currie, K. Cameron, F.M. Dias, H.L. Spencer, E.A. Bayer, C.M. Fontes, S.P. Smith, Z. Jia, Small angle X-ray scattering analysis of *Clostridium thermocellum* cellulosome N-terminal complexes reveals a highly dynamic structure, J. Biol. Chem. 288 (11) (2013) 7978–7985.
[12] M.A. Currie, J.J. Adams, F. Faucher, E.A. Bayer, Z. Jia, S.P. Smith, Scaffoldin conformation and dynamics revealed by a ternary complex from the *Clostridium thermocellum* cellulosome, J. Biol. Chem. 287 (32) (2012) 26953–26961.
[13] M. Hammel, H.P. Fierobe, M. Czjzek, V. Kurkal, J.C. Smith, E.A. Bayer, S. Finet, V. Receveur-Brechot, Structural basis of cellulosome efficiency explored by small angle X-ray scattering, J. Biol. Chem. 280 (46) (2005) 38562–38568.
[14] A.L. Molinier, M. Nouailler, O. Valette, C. Tardif, V. Receveur-Brechot, H.P. Fierobe, Synergy, structure and conformational flexibility of hybrid cellulosomes displaying various inter-cohesins linkers, J. Mol. Biol. 405 (1) (2011) 143–157.
[15] B. Garcia-Alvarez, R. Melero, F.M. Dias, J.A. Prates, C.M. Fontes, S.P. Smith, M.J. Romao, A.L. Carvalho, O. Llorca, Molecular architecture and structural transitions of a *Clostridium thermocellum* mini-cellulosome, J. Mol. Biol. 407 (4) (2011) 571–580.
[16] A. Valbuena, J. Oroz, R. Hervás, A.M. Vera, D. Rodríguez, M. Menéndez, J.I. Sulkowska, M. Cieplak, M. Carrión-Vázquez, On the remarkable mechanostability of scaffoldins and the mechanical clamp motif, Proc. Natl. Acad. Sci. Unit. States Am. 106 (2009) 13791–13796.
[17] C. Schoeler, K.H. Malinowska, R.C. Bernardi, L.F. Milles, M.A. Jobst, E. Durner, W. Ott, D.B. Fried, E.A. Bayer, K. Schulten, H.E. Gaub, M.A. Nash, Ultrastable cellulosome-adhesion complex tightens under load, Nat. Commun. 5 (2014) 5635.
[18] A.M. Vera, M. Carrión Vázquez, Direct indentification of protein-protein interactions by molecular force spectroscopy, Angew. Chem. Int. Ed. 55 (2016) 13970–13973.
[19] A.L. Carvalho, F.M.V. Dias, J.A.M. Prates, T. Nagy, H.J. Gilbert, G.J. Davies, L.M.A. Ferreira, M.J. Ramao, C.M.G.A. Fontes, Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13809–13814.
[20] A.L. Carvalho, F.M. Dias, T. Nagy, J.A. Prates, M.R. Proctor, N. Smith, E.A. Bayer, G.J. Davies, L.M. Ferreira, M.J. Romão, C.M. Fontes, H.J. Gilbert, Evidence for a dual binding mode of dockerin modules to cohesins, Proc. Natl. Acad. Sci. U.S.A. 104 (9) (2007) 3089–3094.
[21] A.M. Nash, S.P. Smith, C.M.G.A. Fontes, E.A. Bayer, Single *versus* dual-binding conformations in cellulosomal cohesin-dockerin complexes, Curr. Opin. Struct. Biol. 40 (2016) 89–96.
[22] J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins, J. Biomol. NMR 20 (2001) 71–75.
[23] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins employing pulsed field gradients, Prog. Nuc. Mag. Res. 34 (1999) 93–158.
[24] D. Pantoja-Uceda, J. Santoro, Amino acid type identification in NMR spectra via β- and γ-carbon edited experiments, J. Magn. Reson. 195 (2) (2008) 187–195.
[25] Goddard, T. D.; Kneller, D. G. SPARKY, 3; San Francisco.
[26] Y.S. Jung, M. Zwackstetter, Mars - robust automatic backbone assignment of proteins, J. Biomol. NMR 30 (1) (2004) 11–23.
[27] C. Camilloni, A. Di Simone, W.F. Vranken, M. Vendruscolo, Determination of secondary structure population in disordered state of proteins using nuclear magnetic resonance spectroscopy, Biochemistry 51 (2012) 2224–2231.
[28] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics, J. Biomol. NMR 24 (2) (2002) 149–154.
[29] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J. Biomol. NMR 44 (2009) 213–223.
[30] M.V. Berjanski, D.S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts, J. Am. Chem. Soc. 127 (43) (2005) 14970–14971.
[31] P. Güntert, Automated NMR structure calculation with CYANA, Meth. Mol. Biol. 278 (2004) 353–378.
[32] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1) (1996) 51-59, 29–32.
[33] V.B. Chen, I.W. Davis, D.C. Richardson, KiNG (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program, Protein Sci. 18 (11) (2009) 2403–2409.
[34] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, Nucleic Acids Res. 42 (WI) (2014) W252–W258.
[35] M. Mompeán, V. Romano, D. Pantoja-Uceda, C. Stuani, F.E. Baralle, E. Buratti, D.V. Laurents, Point mutations in the N-terminal domain of transactive response DNA-binding protein 43 kDa (TDP-43) compromise its stability, dimerization, and functions, J. Biol. Chem. 292 (28) (2017) 11992–12006.
[36] G.A. Tavares, P. Beguim, P.M. Alzari, The crystal structure of a type 1 cohesin domain, J. Mol. Biol. 273 (1997) 701.
[37] A. Chakrabartty, T. Kortemme, R.L. Baldwin, Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions, Protein Sci. 3 (1994) 843–852.
[38] K. Fujiwara, H. Toda, M. Ikeguchi, Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type, BMC Struct. Biol. 12 (2012) 18.
[39] U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski, G. Fischer, Side-chain effects of peptidyl-prolyl *cis/trans* isomerization, J. Mol. Biol. 279 (1998) 449–460.
[40] F. Ramos-Martín, R. Hervás, M. Carrión-Vázquez, D.V. Laurents, NMR spectroscopy reveals a preferred conformation with a defined hydrophobic cluster for polyglutamine binding peptide 1, Arch. Biochem. Biophys. 558 (2014) 104–110.
[41] K. Wüthrich, M. Billeter, W. Braun, Polypeptide secondary structure determination by nuclear magnetic resonance obserbation of short proton-proton distances, J. Mol. Biol. 180 (1984) 715–740.
[42] S.P. Smith, P. Béguin, P.M. Alari, K. Gehring, ^{1}H, ^{13}C, ^{15}N NMR sequence-specific resonance assignment of a *Clostridium thermocellum* type II cohesin module, J. Biomol. NMR 23 (2002) 73–74.
[43] Z. Cui, Y. Li, Y. Xiao, Y. Feng, Q. Cui, Resonance assignments of a cohesin and dockerin domains from *Clostridium acetobutylium* ATCC824, Biomol. NMR 7 (2013) 73–76.
[44] D.L. Minor, P.S. Kim, Measurement of β-sheet forming propensities of amino acids, Nature 367 (1994) 660–663.
[45] D.L. Minor, P.S. Kim, Context is a major determinant of β-sheet propensity, Nature 371 (1994) 264–267.
[46] C.N. Pace, D.V. Laurents, R.E. Erickson, Urea denaturation of barnase: pH dependence and characterization of the unfolded stated, Biochemistry 31 (10) (1992) 2728–3734.
[47] G.J. Gerwig, P. de Waard, J.P. Kamerling, J.F. Vliegenthart, E. Morgenstern, R. Lamed, E.A. Bayer, Novel O-linked carbohydrate chains in the cellulase complex (cellulosome) of *Clostridium thermocellum*. 3-O-methyl_N-acetylglucosamine as a constituent of a glycoprotein, J. Biol. Chem. 264 (2) (1989) 1027–1035.
[48] G.J. Gerwig, J.P. Kamerling, J.F. Vliegenthart, E. Morag, R. Lamed, E.A. Bayer, The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of *Clostridium thermocellum* and *Bacteroids cellulosolvens*, J. Biol. Chem. 268 (36) (1993) 26956–26960.